CMSC 257 Assignment 6: Batch shell script, Regular Expression, and Version Control (git)

Project due date: 11:59 pm EST, 12/11/16

What is a batch shell script?

- shell reads from an input file (You know about this from assignment – 4)

What are regular expressions?

- A regular expression (regex or regexp for short) is a special text string for describing a search pattern. You can think of regular expressions as wildcards on steroids. You are probably familiar with wildcard notations such as *.txt to find all text files in a file manager.

What is version control tool?

- Version control is a system that records changes to a file or set of files over time so that you can recall specific versions later.
- For this project you need an account on GitHub or Bitbucket, both provide version control with Git.
- GitHub provides student developer pack, which will give you access to lot of useful tools and space for 5 private repositories.

    GitHub link -- https://education.github.com/pack
    Bitbucket link -- https://bitbucket.org/product

Instructions:

1. Write a Batch shell script to read a text file.
2. Use regular expression to do the below operations.
    a. Count the number of lines in the file
    b. Count the number of words in the file
    c. Find the most repetitive word in the file.
    d. Find the least repetitive word in the file
    e. Find the word count with below-given characteristics.

      i. Starts with "d" and ends with "d", words can be both upper and lower case. (ex: Demand or demand)

     ii. Starts with "A" or "a" and ends with any letter. (example : America, Amazon)

   iii. Count the numeric words. (example: 53254 or 2513354, but it should not be "hit34jdk").

   iv. Count the alphanumeric words. (example : "abc254fj" or "125gsgd")

3. Push the assignment to private repository on GitHub or Bitbucket

4. You should submit a LastName_FirstName.tar.gz or LastName_FirstName.zip file containing a shell script and report.txt. Report should have a link to GitHub or Bitbucket repository.

Note: Late assignments will lose 5 points per day up to a maximum of 3 days. Code must be submitted in the prescribed format. For questions on grading, contact the TA: Syed Khajamoinuddin. Email: lnusk@vcu.edu

---